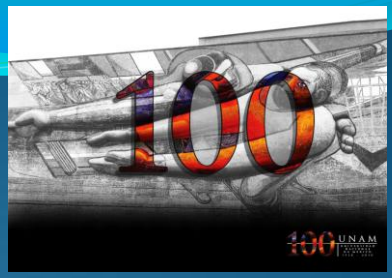
Encuentro de Mexicanistas
Educación, Ciencia y Cultura 2010



"Discutamos México en Europa.
Encuentro de Mexicanistas 2010.
Educación, Ciencia y Cultura"

"DIGITAL PRESERVATION IN MEXICO : AN OVERVIEW"

Belgium. September 20-22, 2010

Juan Voutssas M.

Digital Archives & Information

We have observed that both organizations and individuals all over the world store more and more information every year. Due to the low-cost, minimum space and little effort, there has been a fundamental change in the way we perceive information storage. All kinds of information: texts, photographs, music, video, databases, and so forth, can be stored today in sizes and costs that were just unthinkable two decades ago. This is undoubtedly an advantage and opens great possibilities.

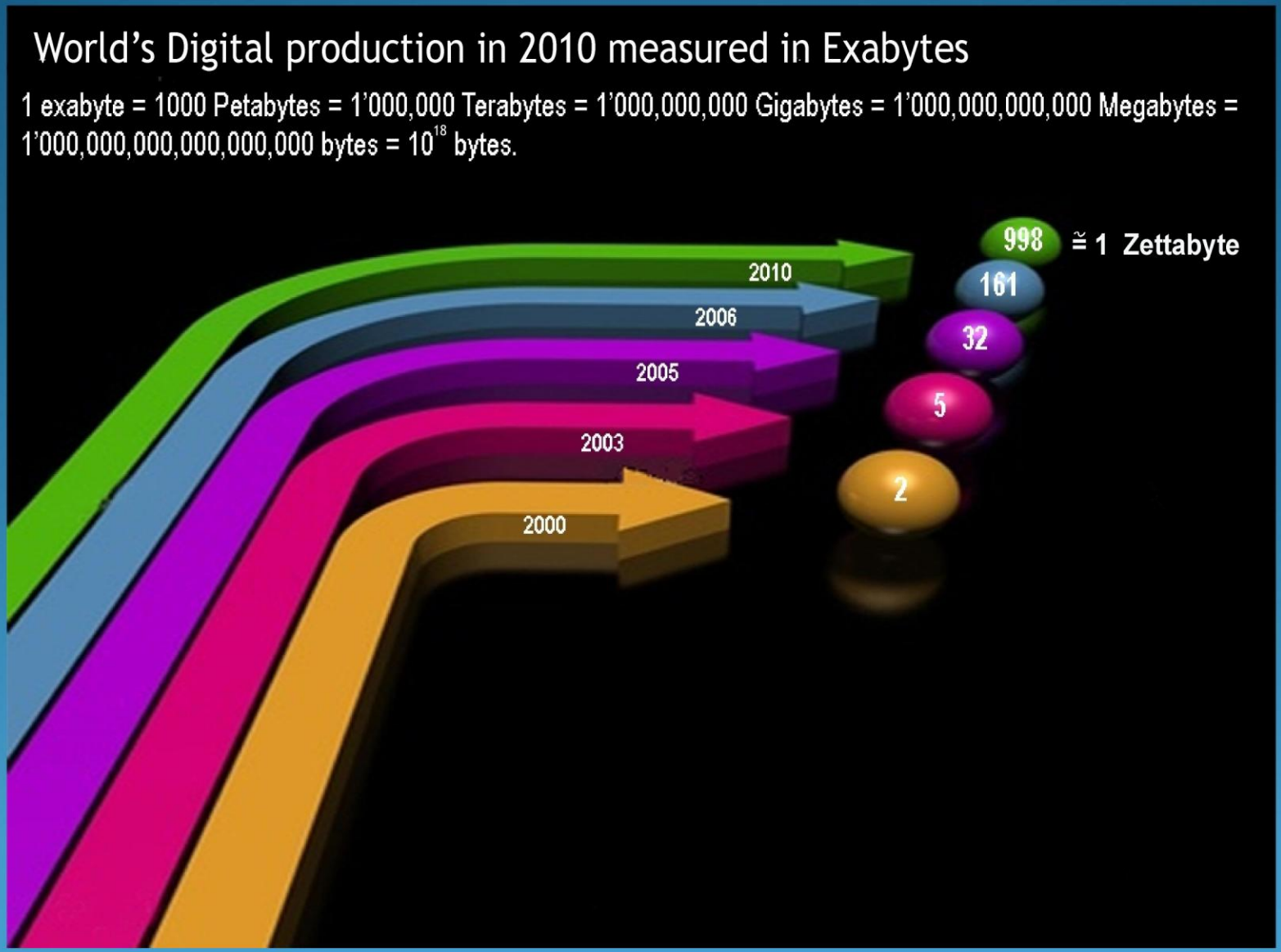
Storing & Preserving

But we often mistake storing for preserving: to store and to preserve are not synonymous. In the same proportion that effortless storage has grown, so does the risk of damage or loss of information and thus, the challenge to preserve it arises. The simple fact of recording some kind of information on a reliable storage device does not mean that it will be available forever. How can we guarantee that the important information we do store today will still be there in -let's say- 20, 100, 200 years? And not only that it will be there; how can we guarantee that, -being there-, it will be readable, trustable and therefore, usable?

Digital Archives

Digital Files & Archives

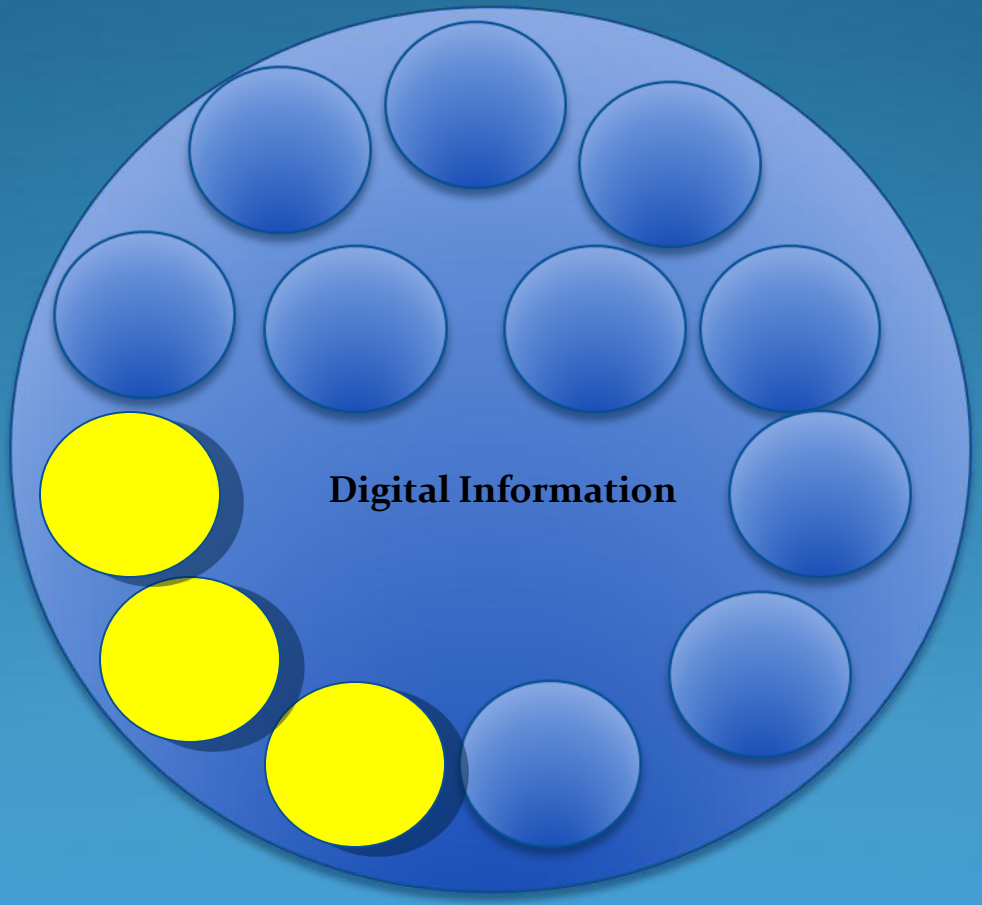
Growth of Digital Information per year



A forecast of Worldwide Information Growth Through 2010: http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf

Avogadro's number, also known as Avogadro's constant, is defined as the quantity of atoms in precisely 12 grams of Carbon 12. The designation is a recognition of Amedeo Avogadro, who in early 19^{th} century was the first to state that a gas' volume is proportional to how many atoms it has. Avogadro's number is given as $6.02214179 \times 10^{23}$ mol^{-1}. It is considered one of the very large numbers in nature and because of that, a boundary.

Growth of Digital Information



Analog Information

Digital Information

Defining the issues

In order to address the issues surrounding long term digital preservation of digital information it is important that we first define certain terms. Digital preservation is *“the specific process of maintaining digital materials during and across different generations of technology over time, irrespective where they reside; this regards the whole of the principles, strategies, policies, rules and proceedings that controls the physical and technological stabilization and protection of the intellectual form of acquired records intended for their continuing, enduring, stable, lasting, uninterrupted and unbroken chain of preservation, without a foreseeable end”*. Long-term is defined as *"a portion of time, a chronological division long enough to be concerned with the impacts of changing technologies, including new operating systems, data and device formats and standards, support for old and new media, or even with the impacts of a changing user community”.*

(www.interpares.org 2006)

Main issues related to this long-term preservation of trustworthy documents

- The immense variety of digital materials to be preserved: books, journals, news and newspapers, photographs, music, films and videos, music scores, scientific publications and databases, maps, films, records and archives.
- Digital records derived from innumerable daily transactions such as: registries, certificates, banking, business, accounting and financial transactions, acts, legal records, mails, reports, contracts and treatises, patents, court sentences, certificates, learning and school records, medical records, insurance, tax payments, human resources, elections, e-mails, web sites and so on. Many of these transactional records will produce historical archives.
- The huge amount of retrospective material still to be digitized: millions of documentary items existing today upon old media within libraries, archives, museums, collections, corporate bodies, offices, governmental agencies, companies, and so forth.
- The ephemeral nature and lose rate of many digital materials; i.e. web pages.
- The lack of awareness about the cultural and historical value of documentary repositories.

Main issues related to this long-term preservation of trustworthy documents (2)

- The technological issues that affect preservation of digital information through time: media, formats, standards, hardware and software, technological obsolescence.
- Information risks and information security.
- Restrictions due to copyrights and information access; regulations about privacy of information.

Existence of suitable metadata which will be useful for future searching and retrieval of information.

Existence of future access and distribution policies and technologies on documentary materials; closing the “digital gap” for future generations.

- The costs associated with preservation. Long term preservation of documentary collections is not cheap, despite what some people think. Although it is cheaper or rather “less expensive” than paper preservation, cost is still an important factor and includes: the costs of digitizing, editing, documenting, storing (this one including cost of security) and updating (migrating, transforming, etc); people usually notice only the storing cost.

The Old Approach

Massive digital information storing –speaking in terms of *"a lotta bytes"*- is relatively recent and in general terms, most of these subjects have not been thoroughly studied or understood. Nevertheless, there exists an urgent and unstoppable pressure within organizations that have to deal with document preservation in digital forms. As noted previously we can distinguish in the literature the already mentioned "technology approach", typical from the eighties and the nineties, a "late 20th century approach" which established that technology was solely responsible for digital materials preservation problems and therefore the answers to this predicament would be exclusively through technology. Now we have developed the current integrative multifactor vision or "21st century approach": complex components of the problem – complex components for the solution.

The New approach

The big framework of the new approach is given by the “factors” involved in long term digital preservation. There are many factors and they are complex. To understand them better, we gathered them into six groups: cultural, technical, legal, methodological, economic and social factors. Each one of these groups represents a particular aspect rarely isolated, each one of them tending to affect one another. Understanding the whole of those factors and not only the technological one, we can deal with a “21st century approach”: complex components of the problem – complex components for the solution.

The New approach

This modern holistic multifactor approach has been developed in developed countries. The previously mentioned initiatives and production of knowledge comes mainly from North America, Europe and Australia. We found no significant research literature or preservation initiatives in Latin American countries, including Mexico.

The problem in Latin America

We can observe that the problem is becoming particularly acute -mostly unnoticed and thus overlooked- in all countries of this region. Most countries are actively working on digital libraries and archives but with little or no study about the long term digital future. A lot of this work is usually done with collections in libraries and archives that have been neglected and have minimum budgets, lack of trained professional personnel, and practically no strategic plans for constructing and preserving digital document collections. In best cases, they are treated with the old approach. With the current ease, low-cost and availability of massive storing devices the region is already creating huge amounts of document collections with little or no awareness of how to deal with their future preservation in a proper way. This is undoubtedly endangering the greatest part of those collections for the future. The existing exceptions just prove the rule. We can find some isolated attempts in some countries of the region: Brazil, Chile, Argentina, Mexico, but none of the size and significance of those in other more developed countries and regions.

The problem in Latin America

Certainly there are in the region projects and initiatives regarding digital libraries, digital archives, digital publishing, information society, and so forth, such as Scielo (Scientific Electronic Library On-Line) in Brazil, e-lac (e-society in Latin America and the Caribbean, but these are digitization and publishing projects, not digital preservation.

The solution

Librarians, archivists and other information professionals have developed handy techniques that can be summed up into several principles that are very helpful while dealing with document preservation. We established seven “principles” to be considered while planning such projects: selection, quality, permanency, accessibility, availability, trustworthiness and functionality. They give us a wide panorama of mostly all the elements that we must have in mind on a daily basis to cope with preservation of digital collections, in a proper and professional way and in a wide variety of contexts and environments.

Planning and acting is important

it is important to emphasize that special efforts should be made by Latin American planners, government officers, library and archive directors, universities, research and other academic institutions in order to reverse this trend and develop suitable strategies and plans within the countries of the region, or develop regional projects and/or agreements about how to start local research on this subject, how to build and propose significant preservation initiatives, how to prepare professional and skilled staff, how to develop and publish regional knowledge about this matter finding local solutions and how to teach government, private organizations and even persons the way to preserve properly in order to revert these harmful effects to their collections and their documental heritage. The projects and solutions developed in other regions and countries of the world are useful guidelines, but they can not just be taken as an “out of the box” solution and implemented “as is” in the countries of this region.

The sooner, the better

The overgrowing production and storage of digital materials and the everyday obsolescence of digital equipment, applications, formats and standards is endangering more and more the future existence of vast amounts of data and information. In many cases, preservation is undertaken by individuals or organizations without knowledge, method and standards rather maintaining instead of preserving the digital materials with little or no degree of authenticity and reliability. Those materials, -if they exist in the future-, will probably be useless because of this. The problem will be acute in the short term -ten or twenty years- but it can be catastrophic in the long term -fifty to several hundred years- if proper measures are not taken now to thwart the possible effects of poor preservation of those digital documents.

"Each renowned author from ancient times whom I recover from oblivion places a new offense and a new cause of dishonor to the charge of preceding generations who, -not satisfied with their own disgraceful shallowness- allowed that other fruitful minds works, the writings that their ancestors produced with great toil and endeavour, to perish through unforgivable neglect. Nothing from their own they had to hand down to those who were to come after, but it would have been enough to forgive them just not to steal posterity of its ancestral heritage".

Francesco Petrarch, ca. 1350.

Thank you

Juan Voutssás

voutssas@unam.mx